ocle
9786297786

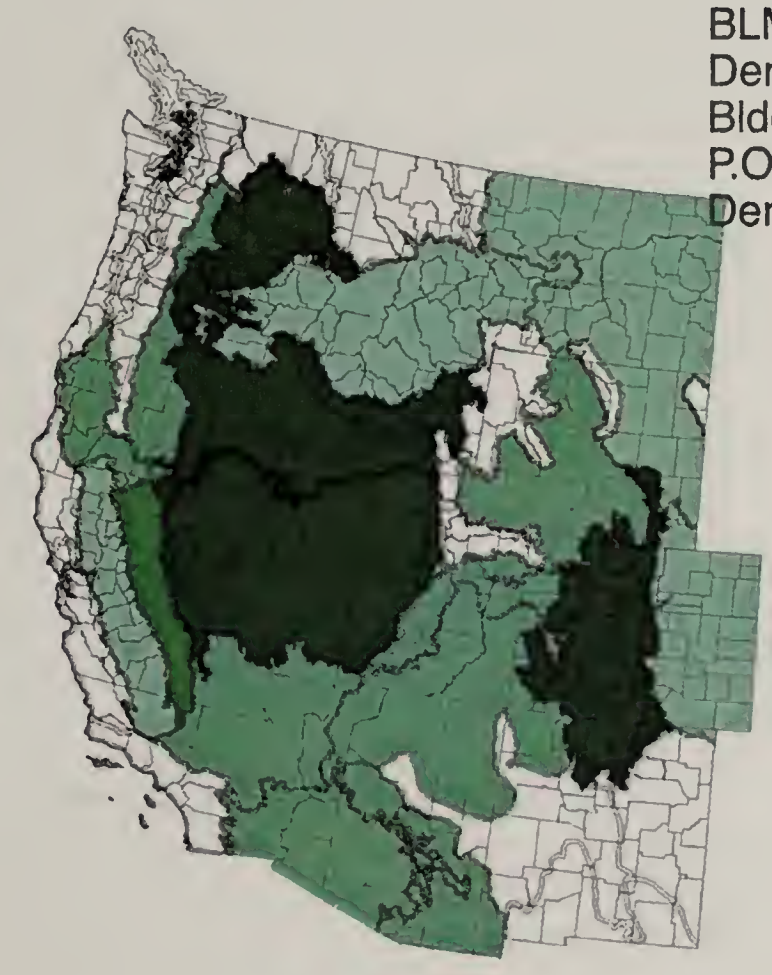
ID
98072927

QK
.911
.F45
2003

A Methodology for Assessing the Ecoregional Distribution of Plant Taxa

M.Q.N. Fellows and L.E. Morse

BLM Library
Denver Federal Center
Bldg. 50, OC-521
P.O. Box 25047
Denver, CO 80225



A NatureServe Report
Prepared for the U.S. Bureau of Land Management

NatureServe

2003

A Methodology for Assessing the Ecoregional Distribution of Plant Taxa

M.Q.N. Fellows and L.E. Morse

A NatureServe Report
Prepared for the U.S. Bureau of Land Management

NatureServe

2003

NatureServe is a non-profit organization
dedicated to providing the scientific knowledge
that forms the basis for effective conservation action.

Citation:
Fellows, M.Q.N., and L.E. Morse. 2003. A methodology for assessing the ecoregional distribution of plant taxa. A report prepared for the U.S. Bureau of Land Management. NatureServe, Arlington, VA.

Cover Image:
Generated in Arc View 3.1, this image shows the ecoregional distribution of the plant species *Ipomopsis tenuituba* in the western United States.

© NatureServe 2003

NatureServe
1101 Wilson Blvd., 15th floor
Arlington, VA 22209

A Methodology for Assessing the Ecoregional Distribution of Plant Taxa

Summary

We created and tested several predictive occurrence models to determine the ecoregional distribution of native plant taxa and applied the preferred model to plants of interest to the Bureau of Land Management, particularly those that may be used in stabilization and rehabilitation activities on federal lands. The models were designed to predict the ecoregions in which plants occurred based on county records for those plants, with particular attention to cases where counties were not wholly encompassed within a single ecoregion. The models were used to place plants into three confidence categories for each ecoregion: presumed or probably present, possibly present, and presumed absent. Across the area we studied, our preferred model placed an additional 6.9 percent of all possible taxon-ecoregion combinations in the highest confidence category compared to limiting this category to plants occurring in counties that were wholly encompassed within a single ecoregion. This was the case even though we chose the preferred model primarily because it was the most conservative (it minimized cases where a taxon determined to be presumed/probable in an ecoregion was actually absent).

This report, and accompanying documentation, complete Task Order #1 to Agreement DLA020218 between the U.S. Bureau of Land Management and NatureServe.

Prepared by NatureServe, 1101 Wilson Blvd., 15th Floor, Arlington, Va. 22209, under the direction of Dr. Larry E. Morse, North American Botanist. Meghan Fellows, Associate Botanist/Information Scientist, was the primary staff researcher. Rob Solomon, Manager, User Support & GIS Services, performed the initial GIS analysis.

NatureServe thanks Dr. John T. Kartesz of Chapel Hill, North Carolina, for permission to use his unpublished county-level species distribution data in this analysis.

Introduction

In order to improve planning, conservation, restoration and management in Bureau of Land Management (BLM) states[1], we created a predictive occurrence model to determine the ecoregional distribution of plant taxa and applied this model to plants of interest to the BLM. The Bureau often uses seed of widespread, common plants for restoration on their lands. They accordingly have an interest in using the most geographically and ecologically appropriate species, subspecies and varieties of native plants, including consideration of whether taxa are native in the particular ecoregion in which they are being considered for planting. BLM provided us with a target list of 77 species, potentially encompassing several hundred infrataxa (subspecies and varieties), that were likely to be used in stabilization and restoration/ rehabilitation efforts on their lands.

Most available floras, checklists and map atlases use political areas, such as counties, states, or nations, to describe the distribution of plants. However, the boundaries of political areas are rarely coincident with the boundaries of biological or ecological features, such as those that are reflected in many types of ecoregion boundaries. Even in cases where the exact biogeography of a plant can be determined from existing sources, doing so taxon-by-taxon is a time-consuming proposition. A predictive model is the most efficient method for attributing taxa to ecoregion(s) where they occur.

We created a predictive occurrence model to determine the ecoregional distribution of plant taxa that places all taxa into one of three categories within each ecoregion: presumed or probably present, possibly present, or presumed absent. We believe this is a simple model that meets the needs of BLM and, more generally, that can be used effectively for large botanical data sets.

At least one predictive model exists in the literature for grouping ecologically similar political units (counties) into more ecologically-based units for similar purposes. Published by the U.S. Department of Agriculture, Forest Service[2] (USFS), the model assigns each county to a "subregion." We have used the same basic approach for our model; however, we have avoided some of the error associated with the USFS model by not forcing a county to represent only one region. Our work indicates that this can be a significant oversimplification. Using The Nature Conservancy's (TNC's) ecoregions[3], for example, we found that only one-third of the counties in the BLM states occur entirely within a single ecoregion. Lane County, Oregon, one of the more ecologically diverse counties in the BLM states, is assigned to only one subregion in the USFS model

---

[1] We refer to those states where the BLM has significant areas of responsibility: Arizona, California, Colorado, Idaho, Montana, Oregon, Nevada, New Mexico, Utah, and Washington.

[2] Rudis, V. A. 1999. Ecological subregion codes by county, coterminous United States. General Technical Report SRS-36. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

[3] Nature Conservancy Ecoregions of the United States, 2001. TNC Ecoregions. Accessed May 2003, Online at ftp://ftp.tnc.org/data/national/usa/tnc_us_eco2001.zip.

"based on the predominant section[4] within that province" even though only 52.1% of the county actually occurs in that section. Lane County, in our model, includes portions of three TNC ecoregions: the Pacific Northwest Coast Ecoregion, the Puget Trough-WillametteValley-Georgia Basin Ecoregion, and the West Cascades Ecoregion. A plant that occurs in Lane County could exist in any or all of the 3 ecoregions; our model weights the likelihood of occurrence in any one ecoregion based on the proportion of the county area occupied by the ecoregion and the total number of counties within the ecoregion in which that plant occurs.

Data and General Methods

The quality and accuracy of the model is directly related to the quality and accuracy of the datasets we used to develop it. For that reason, we used nationally recognized standard datasets including: data in NatureServe's Central Databases[5], The Nature Conservancy Ecoregion Coverage[6] and the U.S. County Coverage[7]. John Kartesz[8] provided the county distribution of the target taxa in draft form. His data are drawn primarily from floristic and monographic literature, integrated to follow a single taxonomic classification.

In general, there are three types of data that are required to produce a model: A calibration dataset that parameterizes the model, a testing dataset that allows a evaluation of the relative accuracy of various models, and the raw dataset that provides the data for the parameters the model includes. We generated the calibration and test datasets from local and statewide floras from Wyoming[9], Colorado[10] and Utah[11]. In selecting taxa for the calibration dataset, we drew from a wide geographic area in order to include a range of variables that are known to affect species occurrence including latitudinal gradients and habitat heterogeneity. We selected taxa from the Wyoming, Colorado, and Utah floras and used habitat and distribution information to determine in which TNC ecoregions the plants occurred. We made an attempt at selecting these taxa randomly; however, it was often unclear from floras alone if a species occurred in a specific ecoregion so the final list of taxa included in the calibration dataset is decidedly opportunistic. In total, we were able to determine ecoregional occurrence for 325 taxa in

---

[4] Section refers to a region that has "similar geomorphology, geologic origin, drainage networks, topography, and regional climate."
[5] NatureServe. 2003. NatureServe's Central Databases. Arlington, VA. U.S.A.
[6] Ibid, footnote 3.
[7] Environmental Systems Research Institute, Inc. (ESRI), 2001. U.S. Counties (Detailed). Accessed May 2003, Online at \\Bioticstest\GIS_data\USA\Contiguous_US\Albers\Political_Boundaries\Counties_100k.shp.
[8] Kartesz, J.T. 2003. County Distribution Data for Vascular Plants of the Conterminous Western United States (draft dataset).
[9] Wyoming Plant Atlas: http://www.esb.utexas.edu/tchumley/wyomap/list.htm#A (accessed August 2003 – 5 September 2003)
[10] Weber, W. A. and R. C. Wittmann. 1996. Colorado Flora Eastern Slope. University Press of Colorado; Weber, W. A. and R. C. Wittmann. 1996. Colorado Flora Western Slope. University Press of Colorado.
[11] Albee, Shultz and Goodrich. 1988. Atlas of the Vascular Plants of Utah. Utah Museum of Natural History.

Wyoming, Colorado and Utah by comparing occurrence data with maps of ecoregion boundaries and from interpreting habitat descriptions in the floras. The test dataset is based on one ecoregion, Wyoming Basins, because the Wyoming plant atlas[10] was detailed enough that we were able to assess presence or absence in the Wyoming Basins for over 170 taxa. Although these calibration and test datasets are limited in scope and therefore not applicable to the entire U.S., they are representative of the coterminous BLM states.

We also developed calculated fields, such as percent area of each county that was within each TNC ecoregion, using the built-in Intersection and Buffer functions in ArcView 3.1. (Based on this, inclusion of an entire county within a single ecoregion occurred 33.8 percent of the time. A greater percentage of counties—45.2 percent—were included in 2 ecoregions, and the remaining 21% of the counties were split between 3 to 5 ecoregions.)

Whenever spatial data at different scales are used concurrently, several known and recognized errors occur that may affect model selection and parameterization. For example, the overall resolution of the final data is only as good as the lowest resolution data. The county boundary spatial layer we used was at a scale of 1:100,000, with the TNC ecoregion map variable but generally coarser. Because of the relatively coarse scale of the ecoregion map, some degree of error is inherent. Additional error can be attributed to the comparison of a political area with a natural area. It is known and accepted that any two maps will have slight variation between boundaries when the maps are overlaid. This problem is exacerbated when one of the boundaries is natural as compared to a political border since mapping of sometimes dynamic landforms (e.g., coastlines or mountain ridges) can create "slivers" of area that fall outside of any defined area. Such areas of counties that fell between ecoregions were examined in ArcView and lumped with the nearest ecoregion. Additionally, many of the thirty-one ecoregions found within the eleven BLM states extend beyond this eleven-state area, and thus beyond the scope of our study. For this reason, taxa that we determined to be absent from an ecoregion may in fact be present in that region in areas outside the political border of the area we studied. For example, the Black Hills Ecoregion is distributed in both Wyoming, a state we examined, and South Dakota, a state that was not included in this analysis; a "presumed absent" result for a taxon in the Black Hills Ecoregion only pertains to the Wyoming portion of the Black Hills Ecoregion.

One error that increased the number of ecoregions that spatially overlapped with a county is known as "boundary fuzz."[12] Actual, on-the-ground boundaries between ecoregions, communities, and habitats are known to be variable in width; however, on a map natural boundaries are symbolized by an infinitely narrow line. Unlike political borders that can be precisely surveyed, ecoregion boundaries cannot be so accurately defined; a combination of factors, including macroclimate, plant associations, soil, and landforms, may be involved and each factor varies along a gradient across the landscape.

---

[12] see *BioScience*, August 2003 for a discussion of problems in determining ecologically meaningful boundaries

Therefore, although ecoregions are specific units in the landscape that are more similar within themselves than they are to other ecoregions, the edges of ecoregions are inherently hard to define. Given the flexibility in the boundary of each individual ecoregion, five caveats are necessary when working with ecoregions boundaries:

1) Ecoregions are not homogenous; most contain habitats, and therefore taxa, that are not typical of the region
2) Ecoregions reflect a compromise among contributing data sets
3) Even though ecoregion boundaries may appear precise, ecoregion boundaries are usually approximate and do not account for ecotones, edge mosaics, and modern human impacts
4) Ecoregion boundaries have an intrinsic width which generally varies along the entire length
5) Ecoregions can include disjunct areas

In our model development, we tested five different ecoregion boundary buffer widths. The number of ecoregions that overlapped counties was, as expected, affected by the choice of widths. In general, however, in comparing the Kartesz (2003) county distribution data with the TNC ecoregion layer, we followed the following rules:

• If a taxon is not reported by Kartesz (2003) from any of the counties that overlap a given ecoregion within the borders of the BLM states, we made the assumption that the taxon was *presumed absent* from that ecoregion. This assumption is probably false in some cases because few if any counties have complete floristic data. In addition, several ecoregions include area that is beyond the borders of the 11 BLM states; these areas outside our study may in fact provide habitat to "presumed absent" taxa. A finding of presumed absent, therefore, does not imply that the taxon is not there, only that it has not been reported from that ecoregion within the BLM states.

• Taxa were *presumed present* for the entire ecoregion when reported by Kartesz (2003) from a county that has no area outside of the ecoregion boundary (e.g., White Pine County, Nevada is wholly within the Great Basin Ecoregion; all 128 taxa reported to occur in White Pine County are presumed present within the Great Basin Ecoregion). This assumption is based on the correct identification (and correct geographical attribution) of the taxon in such wholly contained counties, and may be affected by errors in the baseline floristic information.

• Taxa that are neither presumed absent nor present based on county occurrence alone were further analyzed with a predictive occurrence model (i.e., these were the only taxa addressed by the models).

We evaluated several predictive occurrence models for use in the eleven BLM states. For the purposes of this project, we decided that the preferred model should be a conservative one that limited cases of predicted presence but actual absence within a given ecoregion (i.e., the preferred model would minimize cases of false positives).

We used the same general procedure to create and evaluate each model. All the models assigned each a taxon-ecoregion combination to a group, numbered 1 through 6 in decreasing order of probability that the taxon occurs in that ecoregion. These groups are qualitative and did not represent the same value across all models, nor do these categories have a linear relationship: Group 1 does not represent a simple factor of Group 6. We ran the test dataset (the 170 taxa for which we believe we could independently assess with a high degree of confidence presence or absence in the Wyoming Basins Ecoregion) through each model and checked the accuracy of each model to determine the preferred one.

The models included a range of combinations of raw parameters and calculated values. Thresholds of the parameters are directly related to threshold values in the calibration dataset (the Wyoming, Colorado, and Utah taxa for which we had independently determined ecoregional occurrence) . For example, we graphed the percent area of an ecoregion comprised by all the counties in which the taxon is reported to occur in that ecoregion, and for the taxa known to be absent from that ecoregion (Figure 1). We then examined the graphs for discontinuities. Taxa that occurred in counties comprising more than 55 percent of the area of any ecoregion were always actually present in that ecoregion; therefore we used 55 percent as the bottom threshold of our Group 2 (the choice of Group 2 as opposed to Group 1 was intended to make the models' predictions more conservative). Likewise, maximum thresholds could be determined from the graphs. No taxa that were actually present in the ecoregion has a "percent area of an ecoregion" less than 5.4 percent; therefore, we made the maximum threshold for Group 6 5.4 percent. Additional characteristics of the graphs, such as averages and standard deviations, were also used to determine thresholds for parameters in the various models (for examples, see the section on other models considered on page 13).

To minimize the effect of boundary or mapping errors resulting from comparing two maps, we created five different ecoregion boundary buffer widths and used these different widths in the various models and examined their effects on the models' accuracy. The width tolerances affected the number of ecoregions that overlapped the 413 counties that were in the eleven BLM states. The default width (infinitely narrow) includes the greatest number of ecoregions within counties. At the next most conservative width of ½-mile from the mapped line, at least one ecoregion was excluded from 5 different counties. The most conservative ecoregion boundary width tested (5 miles) excluded one or more ecoregions from 60 counties. We also analyzed widths of 1.5 and 3 miles.

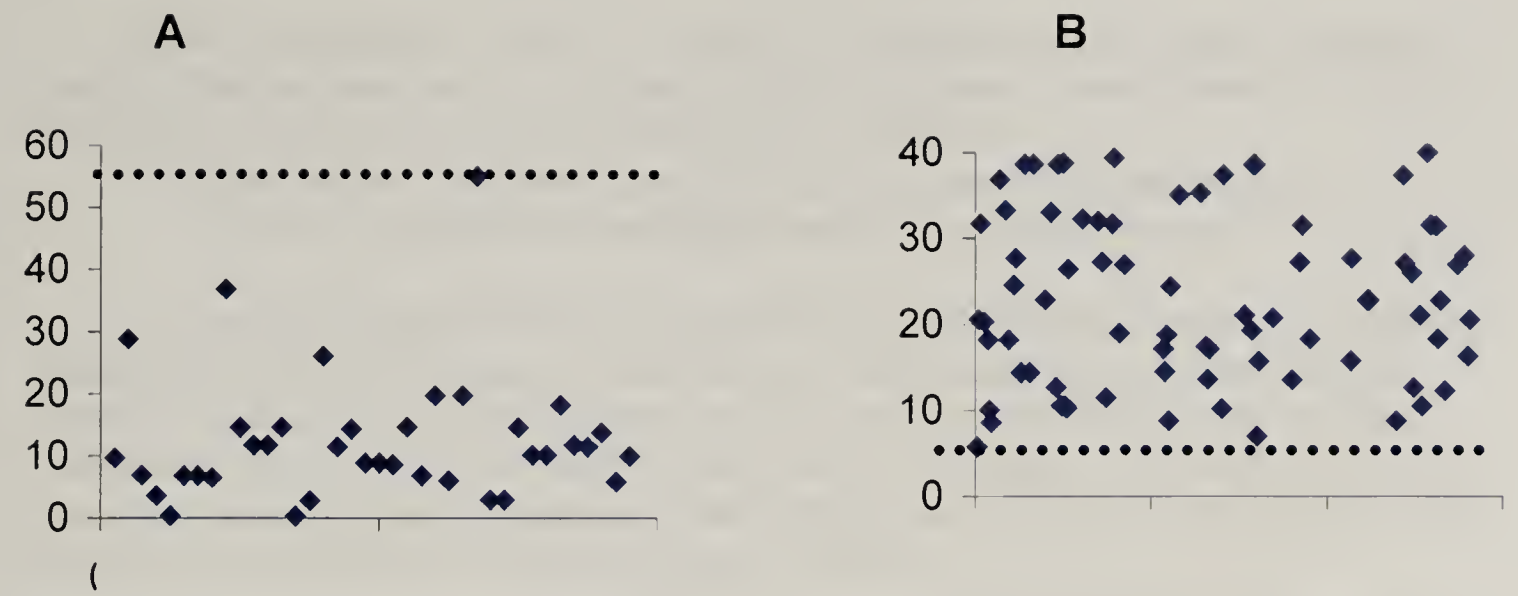


Figure 1. Examples of types of graphs used for determining threshold rules. The percent area of an ecoregion represented by all the counties in which a taxon is reported to occur is shown on the vertical axis. The data points represent taxa that are known to be absent in graph A and taxa that are known to be present in graph B. The threshold in graph A, represented by the dotted line, suggests a minimum value for a very high probability (either Group 1 or Group 2) would be percent area of greater than or equal to 55%. Graph B suggests an upper threshold of 5.4% for a very low predicted occurrence, either Group 5 or Group 6.

Results and Discussion

After comparisons with the test dataset, we designated the Weighted Average model as the preferred one[13]. The premise of the Weighted Average model is that the probability that a taxon occurs in an ecoregion is related to:

- The number of counties in which a taxon occurs within the ecoregion
- The total area within an ecoregion that those counties represent.

Total area is determined by summing the area of overlap of a county or counties within the ecoregion of interest. While two taxa might each occur in 15 of the 28 counties within an ecoregion, if the specific 15 counties differ between the two, the counties could occupy very different areas within the ecoregion (Figure 2). Basically, a taxon with a county distribution that includes more land area within an ecoregion will have a greater probability of occurring there (Table 1).

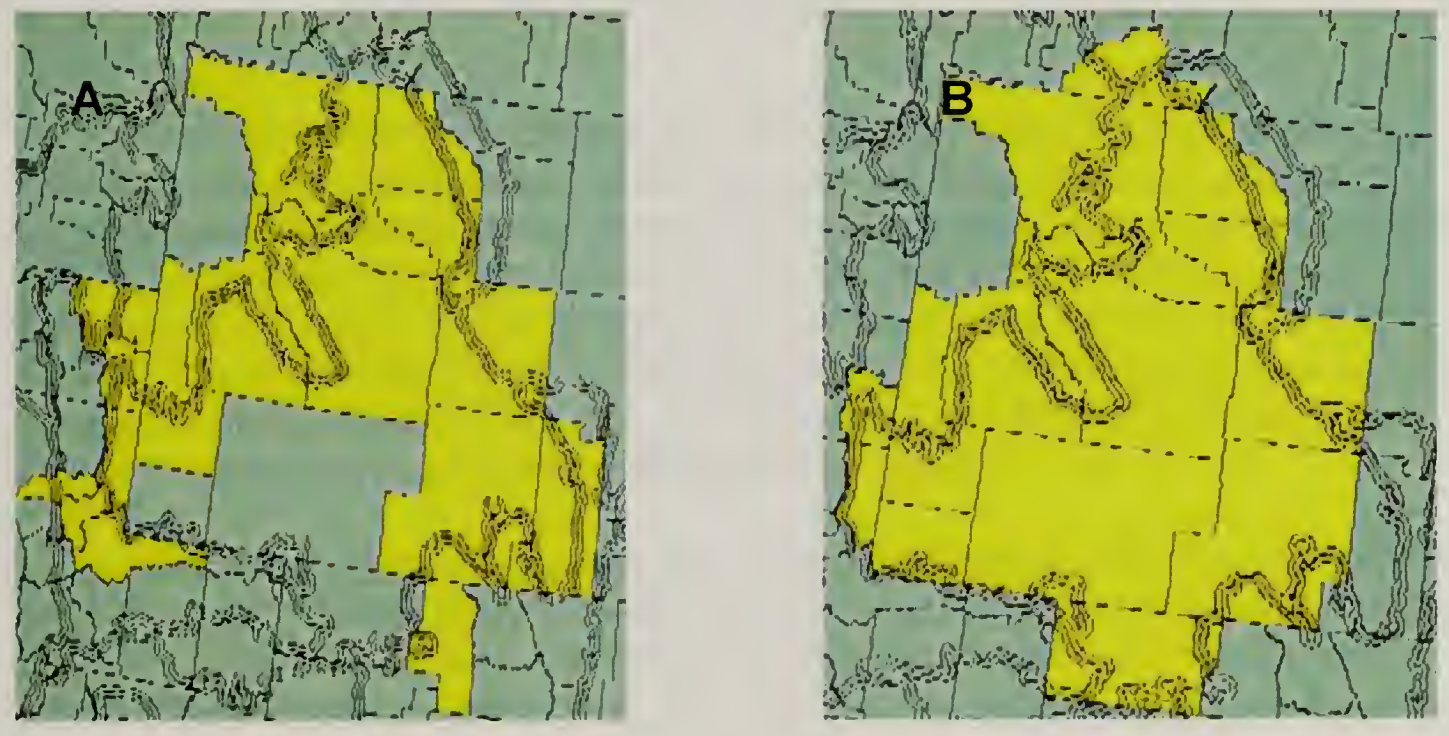


Figure 2. Example distributions of two imaginary taxa, A and B. Highlighted counties represent taxon presence. Dashed lines represent county borders; the triple solid line represents an ecoregion boundary with a finite width. Both taxa occur in the same number of counties within the ecoregion. However B occurs in counties that comprise a greater area (86%) of this ecoregion than does A, which occurs in counties that comprise only 60% of the ecoregion. This suggests that B has a greater probability of occurring in this ecoregion that A.

To mitigate for the errors associated with boundary fuzz, in applying this model we used the most conservative 5-mile ecoregion boundary. Counties that were in an ecoregion *only* within the 5-mile boundary were still considered part of that ecoregion but were weighted less (see Table 1). Counties that became wholly contained within an ecoregion when the 5-mile boundaries were used were accommodated in the model the same as partial counties.

---

[13] For a discussion of other models considered, see page 13.

Table 1. Steps for assigning weight to counties where plants occur, based on threshold characteristics determined from the calibration dataset.

Step	Action
1	If the county is in the ecoregion but exclusively within the 5-mile ecoregion boundary, weight county 0.1
2	If the county or counties total area within the ecoregion is >= 67% of the area of the ecoregion, weight each (excluding those weighted in Step 1) county 1.5[14]
3	All other counties weight = 1
4	Normalize sum of species/county/ecoregion combinations with total number of counties within an ecoregion (take the sum of weights and divide by total number of counties in the ecoregion).

Once we assigned the weights, we summed them, which produced a value between 0 and 100. To "translate" these values into our six Groups, we graphed the values produced by the calibration dataset and looked at the average, standard deviation, and discontinuities in the graph for taxa known to be present or absent in that set. Examining that graph helped us to assign the most appropriate number ranges to each Group. (The numbers produced by summing the weights in other models we tried were "translated" into groups the same way, but each model was different in terms of the exact range of numbers that were placed into a particular Group.)

The Weighted Average model was our preferred model in large part because, in comparison with the test dataset, it had a low rate of false positives (0%) as determined from the number of taxa that the model placed in Groups 1 and 2 that were actually absent (Table 2).

Table 2. Number of taxa placed in each Group by the Weighted Average model that were determined to be actually present or absent from the Wyoming Basins ecoregion. None of the 81 taxa that the model placed in Groups 1 or 2 are actually presumed to be absent from the ecoregion.

Group	**1**	**2**	**3**	**4**	**5**	**6**
Actually Present	**57**	**24**	24	46	0	3
Actually Absent	**0**	**0**	2	15	0	1

[14] Sixty-seven percent was used instead of fifty-five percent (as shown in Figure 1) because the use of the 5 mile wide ecoregion boundary changed the threshold value for this parameter.

The specific weights assigned to counties in the Weighted Average model was determined through trial and error (comparisons with the test dataset). Varying the weights given to a county changes the results of the model (see Tables 3 and 4), but the rate of false positives is under 1.5% for most reasonable permutations.

Table 3. Steps for assigning alternative weights to counties where plants occur (one example of alternative weights tried).

Step	Action
1	If the county is in the ecoregion but exclusively within the 5-mile ecoregion zone, weight county 0.1
2	If the county or counties total area within the ecoregion is >= 67% (excluding those weighted in Step 1) of the area of the ecoregion, weight each county 1.9
3	All other counties, weight = 1
4	Normalize sum of species/county/ecoregion combinations with total number of counties within an ecoregion.

Table 4. Number of taxa placed in each Group by the Weighted Average model (using alternative weights) that were determined to be actually present or absent from the Wyoming Basins. One of the 91 taxa that this permutation of the model placed in Groups 1 & 2 (1.0%) are actually presumed to be absent from the ecoregion..

Group	1	2	3	4	5	6
Actually Present	**83**	**7**	23	25	16	0
Actually Absent	**1**	**0**	5	3	9	0

In addition to minimizing false positives, the Weighted Average model was selected as the preferred method because it correctly predicted the presence of taxa (based on Groups 1 and 2) in the test dataset 99-100% of the time, depending on specific thresholds used in the application of the model. In addition, the Weighted Average model was also relatively successful at predicting presence when the total area comprised by the counties of occurrence is relatively small. However, the Weighted Average model was the most conservative, assigning only 81 taxa to the two highest predictive occurrence categories.

Before applying the model to the target list of species provided by BLM, we first determined that a total of 432 species, subspecies, and varieties are associated with the targeted 77 species. Of these taxa and infrataxa, 409 occurred somewhere within the

BLM states. Based on a simple overlay of counties where these taxa occur and ecoregions, taxa *could* occur in 5,622 of the possible 12,679 taxon-ecoregion combinations (409 taxa * 31 ecoregions). Inclusion of an entire county within a single ecoregion occurred 33.8 percent of the time, allowing the immediate assignment of 40.7 percent of the taxon-ecoregion pairs at the level of "presumed present" in the model. Of the 3,334 additional taxon-ecoregion combinations, a further 6.9 percent were assigned to "presumed or probably present" based on the results of the predictive model. Across all ecoregions, this model predicted as "possibly present" 52.4 percent of the taxa known to occur in western ecoregions (Table 5). Which groups were placed in the "presumed or probably present" vs. "possibly present" categories ultimately depends on the users' tolerance for error. If Group 3 were also considered "presumed or probably present," it would add only a minimal amount of additional error (2 percent).

A spreadsheet with the complete results of the Weighted Average model, showing which taxa are predicted to occur in which ecoregions, has been provided to BLM along with this report.

We caution that this model should only be used for the 11 BLM states because the calibration and test datasets were drawn from this region. A model for a different region (e.g., the New England states, the Gulf states) should have both calibration and test datasets selected from the area to be modeled. Furthermore, the calibration and test datasets should ideally be randomly, not opportunistically, derived—a goal that would probably necessitate field truthing. However, if both the calibration and test datasets are randomly selected from the entire range, creating independent samples, a statistical probability can be assigned to the various predicted occurrence categories.

An additional caution is that the Kartesz county dataset records floristic presence, without distinguishing native from non-native status at the county level. Therefore, if a taxon is escaped from cultivation, or otherwise suspected to have spread into an ecoregion due to human actions, our results should not be taken to indicate that it is native there. For example, some of the native varieties of the yarrow *(Achillea millefolium)* may be present in some ecoregions only as a direct or indirect result of human actions. However, for the particular taxa in the present study, such cases may be rare.

Although such statistical rigor was beyond the scope of this project, we believe the basic approach we developed here offers a simple and efficient way to predict ecoregional distribution. The approach is neither overly simplistic nor does it rely on detailed, time-consuming research on all individual taxa. As such, this type of approach can be applied, in combination with county occurrence data, to any large botanical dataset. Because of the BLM's specific needs, our preferred model minimizes the risk of over-predicting distribution. Others, with other goals in mind, may chose a somewhat different 'preferred' model.

Table 5. Assignment of taxa to ecoregions based on the Weighted Average model (in bold). Numbers in the Presumed Present and Presumed Absent columns represent the number of taxa automatically assigned to that ecoregion based on unambiguous county distribution in or out of the ecoregion .

	Presumed or probably present			Possibly present				Presumed absent
Ecoregion Name	Presumed Present	1	2	3	4	5	6	Presumed Absent
Apache Highlands	29	2	4	6	124	11	16	221
Arizona-New Mexico Mountains	91		1	2	82		44	193
Black Hills		91			14			308
California Central Coast	60				31		33	289
California North Coast	72				44		1	296
California South Coast	32	10		16	73		13	269
Canadian Rocky Mountains	112				42		21	238
Central Shortgrass Prairie	92				38	12	23	248
Chihuahuan Desert	46			2	68		2	295
Colorado Plateau	148	1	2	2	67	24	28	141
Columbia Plateau	165				41	3	51	153
Fescue-Mixed Grass Prairie		11		41	51			310
Great Basin	173				44	23	31	142
Great Central Valley	21	2	3	8	90	16	37	236
Klamath Mountains	84	2	1	2	85			239
Middle Rockies – Blue Mountains	197				5		30	181
Modoc Plateau and East Cascades		72	3	22	95			221
Mojave Desert	125	3	1	3	100			181
North Cascades and Pacific Ranges		24		21	92			276
Northern Great Plains Steppe	131				7	4	32	239
Okanagan	73				57		16	267
Pacific Northwest Coast	49				34		26	304
Puget Trough - Willamette Valley - Georgia Basin	60				38		22	293
Sierra Nevada		56	4	13	107		20	213
Sonoran Desert	17			9	115	49	7	216
Southern Rocky Mountains	164				30		38	181
Southern Shortgrass Prairie	53				46	40		274
Utah High Plateaus	112	7	1	4	97		28	164
Utah-Wyoming Rocky Mountains	123		1	6	66	7	42	168
West Cascades	59			1	85	24		244
Wyoming Basins		57	27	29	97		22	181
Percent of all possible taxon-ecoregion combinations	40.7	6.0	0.9	3.3	35.0	3.8	10.4	

Other Models Considered

Rules-Based model:

The Rules-Based model independently determines a predicted outcome for two parameters: number of counties in which the taxon occurs, and percent area occupied by those counties within the ecoregion (see Table 6). Based on the calibration dataset, whenever a taxon occurred in more than 20 counties within an ecoregion, it was always actually present in the ecoregion; therefore, we created a threshold at 20 counties and assigned all taxa that occurred in more than 20 counties a "high" predicted occurrence. Similarly, no taxa were actually absent from an ecoregion when the total area of the counties of occurrence within that ecoregion was greater than 82 percent. For that reason, we assigned 82 percent as a threshold for a second parameter predicting the occurrence at "high." Different combinations of the two parameters were assigned to one of the 6 groups. For example, "high" in both parameters meant that we placed the taxon in Group 1 for that ecoregion. This model lead us to place a relatively large number of taxa in Groups 1 and 2 and maximized the number of "true positives" (i.e., taxa correctly predicted to occur in the ecoregion). However, the percent error rate for false positives, 4.5%, was also very high compared to the other models tested (see Table 7).

Table 6. Thresholds and logic for Rules-Based Model.

Step	Action
1	If occurs in > 20 counties in an ecoregion = "High"; (based on the calibration dataset showing no actual absents when taxa occurred in > 20 counties in an ecoregion)
2	If occurs in >= 15 And =< 20 counties in an ecoregion = "High-Mid" (represents 2 standard deviations above the average number of counties occupied by taxa that were actually absent in the calibration dataset).
3	If occurs in <=6 counties in an ecoregion = Low (average for taxa actually absent)
4	If occurs in counties representing >= 82% of area of ecoregion = High (based on the calibration dataset showing no actual absents for cases greater than this)
5	If occurs in counties >= 58% And <82% of area of ecoregion = High-Mid (represents 2 standard deviations above the average for taxa actually absent in the calibration dataset)
6	If occurs in <= 16% of area in ecoregion = Low (average for taxa actually absent)
7	All others = Mid

Table 6. Thresholds and logic for Rules-Based Model (continued)

Group 1	High/High
Group 2	High/High-Mid & High-Mid/High-Mid
Group 3	Mid/High
Group 4	Mid/High-Mid, Mid/Mid, Low/High & Low/High-Mid
Group 5	Low/Mid
Group 6	Low/Low

Table 7. Number of taxa placed in each Group by the Rules-Based model that were determined to be actually present or absent from the Wyoming Basins ecoregion. Five of the 109 taxa the model placed in Groups 1 & 2 (4.5%) are actually presumed to be absent from the ecoregion.

Group	**1**	**2**	**3**	**4**	**5**	**6**
Actually Present	**71**	**33**	0	26	17	7
Actually Absent	**0**	**5**	0	3	4	6

Single factor models, such as Percent Maximum Area Possibly Occupied and Percent Total Counties Present, were also evaluated for predictive ability:

Percent Maximum Area Possibly Occupied:

We calculated Percent Maximum Area Possibly Occupied by summing the county acreage within an ecoregion for all counties for which a taxon is present, divided by the acreage of the entire ecoregion. For example, if a taxon occurs in two counties, one with 500 acres and one with 1500 acres within a specified ecoregion, it would occur in 20 percent of the maximum area possibly occupied if the ecoregion covers 10,000 acres total. The percent maximum area possibly occupied was compared with the calibration dataset for group assignments.

Group 1	>=90%	Group 4	<49 and >= 32%
Group 2	<90 and >= 69%	Group 5	<32 and >= 3.4%
Group 3	<69 and >= 49%	Group 6	<3.4%

This model had the highest number of taxa predicted to occur in Groups 1 and 2 (106) and a low rate of false positives (1.9 percent). It is a possible alternative to the preferred Weighted Average model.

Table 8. Number of taxa plaeed in eaeh Group by the Pereent Maximum Area Possibly Oeeupied model that were determined to be aetually present or absent from the Wyoming Basins eeoregion. Two of the 106 taxa the model plaeed in Groups 1 & 2 (1.9%) are aetually presumed to be absent from the eeoregion.

Group	**1**	**2**	**3**	**4**	**5**	**6**
Aetually Present	**75**	**31**	14	11	21	2
Actually Absent	**0**	**2**	3	4	8	1

Percent Total Counties Present:

We caleulated Percent Total Counties Present by eounting the number of counties for whieh a taxon was present within an ecoregion divided by the total number of eounties within that eeoregion. The result was eompared with the calibration dataset for group assignments.

Group 1	>=95%	Group 4	<40 and >= 14%
Group 2	<95 and >= 59%	Group 5	<14 and >= 5.7%
Group 3	<59 and >= 40%	Group 6	<5.7%

This model had 2.1 percent false positives.

Table 9. Number of taxa plaeed in eaeh Group by the Pereent Total Counties Present model that were determined to be actually present or absent from the Wyoming Basins eeoregion. Two of the 95 taxa the model plaeed in Groups 1 & 2 (2.1%) are aetually presumed to be absent from the eeoregion.

Group	**1**	**2**	**3**	**4**	**5**	**6**
Aetually Present	**26**	**67**	20	33	6	2
Aetually Absent	**0**	**2**	4	8	3	1

We evaluated several other models, mostly combinations and permutations of the above models that altered threshold tolerances or added additional parameters like species diversity. For example, we used a premise that a taxon that was absent from a county with high species diversity (more than 1,730 speeies) is less likely to be in that eeoregion. The model that used this premise was only slightly less effective than the Weighted Average model, but very dependent on full floras being known in order to determine loeal speeies diversity. Weighting counties based on a taxon's presence in a county with low species diversity (fewer than 225 species) is the logical antithesis of the previous method. Using this premise had no effeet on the results of most models in our study sinee the few very low diversity counties happened to be wholly contained within a single eeoregion. In addition, we found that seale dependent parameters, such as a raw

count of counties were not good predictors because of extreme differences in the number of counties per ecoregion (e.g., the Black Hills includes only 2 counties in the BLM states, while Wyoming Basins includes 28). Similarly, we found that using a log-based threshold for area of occurrence was not valuable because this did not account for the large actual differences in areas available within ecoregions.

BLM Library
Denver Federal Center
Bldg. 50, OC-521
P.O. Box 25047
Denver, CO 80225